

Data Management in Clinical Research

Syed Shahnawaz Quadri, Syeda Saniya Iftiqar, Syed Shafa Raoof

Student at Clinisol Research, Hyderabad, Telangana, India

ABSTRACT

A crucial stage in clinical research is clinical data management (CDM), which produces high-quality, reliable, and statistically sound data from clinical trials. This results in a significantly shorter period of time between drug development and marketing. Team members of CDM are laboriously involved in all stages of clinical trials right from commencement to completion. They should be able to sustain the quality standards set by CDM processes by having sufficient process expertise. Colorful procedures in CDM including Case Report Form (CRF) designing, CRF reflection, database designing, data-entry, data confirmation, distinction operation, medical coding, data birth, and database locking are assessed for quality at regular intervals during a trial. In the present script, there's an increased demand to ameliorate the CDM norms to meet the nonsupervisory conditions and stay ahead of the competition by means of brisk commercialization of products. With the perpetration of nonsupervisory biddable data operation tools, the CDM platoon can meet these demands. Also, it's getting obligatory for companies to submit the data electronically. CDM professionals should meet applicable prospects and set norms for data quality and also have the drive to acclimatize to the fleetly changing technology. This composition highlights the processes involved and provides the anthology an overview of the tools and norms espoused as well as the places and liabilities in CDM.

KEYWORDS: *Clinical data cloverleaf norms institute, clinical data operation systems, data operation, e-CRF, good clinical data operation practices, confirmation*

INTRODUCTION:

A clinical trial aims to answer the research question by producing data that can be used to support or refute a theory. The quality of the generated data has a big impact on how the study turns out. What is Clinical Data Management (CDM) and why is it important is the subject under discussion. It is one that research students frequently ask. A clinical trial's relevant and significant component is clinical data management. All researchers, whether consciously or unconsciously, engage in CDM activities throughout their research. In the course of conducting our research, we use a few CDM processes without mentioning the technical steps. This article outlines the CDM procedures and provides the reader with an overview of clinical trial data management.

The process of gathering, scrubbing, and managing subject data in accordance with legal requirements is known as CDM. In order to produce high-quality

How to cite this paper: Syed Shahnawaz Quadri | Syeda Saniya Iftiqar | Syed Shafa Raoof "Data Management in Clinical Research" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-7 | Issue-2, April 2023, pp.485-490, URL: www.ijtsrd.com/papers/ijtsrd55050.pdf



Copyright © 2023 by author (s) and International Journal of Trend in Scientific Research and Development Journal. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0) (<http://creativecommons.org/licenses/by/4.0>)



data, CDM techniques minimize mistakes and missing data while collecting as much data as is practical for analysis. [1] Best practices are applied to ensure that the data is accurate, reliable, and managed appropriately in to achieve this goal. The introduction of software programs that keep an audit trail and make it simple to identify and fix data anomalies has made this possible. Advanced innovations [2] have made it possible for CDM to manage sizable trials and guarantee data quality even in challenging trials. What exactly does "high-quality" data mean? Statistical analysis should be possible with high-quality data, which must be 100 percent accurate. These should conform to the protocol's criteria and the parameters it specifies. This suggests that if there is a deviation or a patient doesn't fit the procedure requirements, we may consider excluding them from the final database. It should be kept in mind that

regulatory agencies might be interested in looking at such data in some circumstances. In a similar vein, clinical researchers are concerned about missing data.

It should be kept in mind that regulatory agencies might be interested in looking at such data in some circumstances. Similarly to this, clinical researchers are likewise concerned about missing data. Data of high quality should have few if any, misses. But most crucially, good data should only have an arbitrarily "acceptable level of variation" that doesn't change the study's statistical analysis's result. Additionally, the data must comply with all applicable regulatory criteria for data quality.

Tools for CDM

There are numerous software programs available for data management; these are referred to as clinical data management systems (CDMS). A CDMS has become necessary in multicenter trials to manage the massive volume of data. There are a few open-source tools available, but most CDMS used by pharmaceutical companies are commercial products. ORACLE CLINICAL, CLINTRIAL, MACRO, RAVE, and eClinical Suite are examples of frequently used CDM tools. These software tools are functionally more or less identical, and neither system has a clear advantage over the other. These software products are pricey and require a high-end IT infrastructure to operate. Additionally, a few global pharmaceutical behemoths use CDMS solutions that are specifically tailored to their operating requirements. Among open-source software, the most well-known ones are PhOSCo, openCDMS, TrialDB, and OpenClinica. These CDM programs are available for free download and offer similar functionality to their paid version. You can download this open-source software from the websites for each of them.

Maintaining an audit record of data management actions is crucial for regulatory submission studies. These CDM tools make sure that the audit trail is preserved and aid in the handling of discrepancies. Several user IDs may be generated with access restricted to data entry, medical coding, database design, or quality check, depending on the roles and responsibilities (described below). This makes sure that each user may only access the functions that are exclusive to their user ID and cannot alter the database in any other way. For tasks that allow data modifications, the software will keep a record of the change made, the user ID of the user who made the change, and the time and date of the change for audit purposes (audit trail). The auditors can determine that no unlawful or misleading changes were made during a regulatory audit by examining the discrepancy

management procedure, the changes made, and the changes themselves.

Regulations, Guidelines, and Standards in CDM

CDM, like other areas of clinical research, has guidelines and standards that must be met. Because the pharmaceutical industry relies on electronically captured data to evaluate medicines, good CDM practices and electronic data capture standards must be followed. These electronic records must adhere to 21 CFR Part 11 of the Code of Federal Regulations (CFR). This regulation applies to electronic records created, modified, maintained, archived, retrieved, or transmitted. This necessitates the use of validated systems to ensure data accuracy, dependability, consistency, and the usage of safe, computer-generated data are all important. time-stamped audit trails to separately document the date and time of operator inputs and actions that add, change, or remove electronic records. th[3] Adequate procedures and controls should be in place to ensure data integrity, authenticity, and confidentiality. If data must be submitted to regulatory authorities, it must be entered and processed in 21 CFR part 11-compliant systems. Most CDM systems are like this, and pharmaceutical companies as well as contract research organisations ensure compliance.

The Good Clinical Data Management Practices (GCDMP) guidelines, published by the Society for Clinical Data Management (SCDM), are a document that provides the standards of good practice within CDM. GCDMP was first published in September 2000 and has since undergone several revisions. The current GCDMP document is the July 2009 version. GCDMP provides guidance on accepted CDM practices that are consistent with regulatory requirements. It covers the CDM process in 20 chapters, highlighting minimum standards and best practices.

The Clinical Data Interchange Standards Consortium (CDISC), a non-profit multidisciplinary organisation, has developed standards to support the acquisition, exchange, submission, and archival of clinical research data and metadata. Metadata is information about the data that has been entered. This includes information about the person who made the clinical data entry or change, the date and time of entry/change, and the details of the changes made The Study Data Tabulation Model Implementation Guidance for Human Clinical Trials is one of the most crucial requirements. (SDTMIG) and the Clinical Data Acquisition Standards Harmonization (CDASH) standards, both of which are freely available on the CDISC website (www.cdisc.org).

The SDTMIG standard [4] describes the details of the data model and standard terminologies and serves as a guide for the organisation. CDASH v 1.1[5] defines the fundamental standards for data collection in a clinical trial and enumerates the basic data information required from a clinical, regulatory, and scientific standpoint.

CDM Procedure

Like a clinical trial, the CDM procedure starts with the end in mind. This shows that the deliverable was taken into account at every stage of the procedure. The CDM method aims to provide a database that is legitimate, error-free, and statistically sound, similar to how a clinical trial aims to provide an answer to a research question. The CDM process begins before the study protocol is finalised in order to achieve this goal.

Review and finalization of study documents

For consistency and clarity, the protocol is examined from the standpoint of database design. The CDM staff will specify the data items to be gathered and the frequency of collection in relation to the visit schedule during this review. The CDM team creates a

Case Report Form (CRF) as this is the first step in converting the protocol-specific actions into data generation. The data fields should have uniform definitions across the board. The CRF should make clear the type of data that needs to be entered. If weight, for instance, must be recorded to two decimal places, Two data boxes should be positioned following the decimal in the data entry field, as shown in Figure 1. The units in which measurements must be performed should also be noted next to the data field, in a similar manner. The CRF needs to be brief, clear, and simple to use (unless you are the one entering data into the CRF). The CRF Completion Instructions should be given to research investigators together with the CRF in order to ensure error-free data collection. The variable is named using either internal naming standards or the SDTMIG when CRF annotation is performed. Annotations are coded terms that are used in CDM tools to denote the study's variables. Figure 1 shows an example of an annotated CRF. In inquiries requiring discrete All possible options will be suitably coded as value options (such as the variable gender containing values for male and female as answers).

Figure 1

DCM = DM[S1]

DEMOGRAPHY	
BRTHDTC [DT] Date of Birth (dd/mm/yyyy): <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	BRTH [C] DVG = YNNA[S1] <input type="checkbox"/> ₁ If Date of Birth unknown
SEX [C] DVG = SEX[S1] Gender: Male <input type="checkbox"/> ₁ Female <input type="checkbox"/> ₂	HEIGHT [N] Height (cm): <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
	WEIGHT [N] Weight (Kgs): <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>

A Data Management Plan (DMP) is created based on these. The DMP document is a road map for dealing with data under foreseeable circumstances and describes the CDM activities that will be carried out during the trial. Table 1 contains a list of CDM activities. The database design, data entry and data tracking guidelines, quality control measures, SAE reconciliation guidelines, discrepancy management, data transfer/extraction, and database locking guidelines are all described in the DMP. Along with the DMP, a Data Validation Plan (DVP) containing all edit-checks to be performed as well as the calculations for derived variables is prepared. The edit check programmes in the DVP aid in data cleaning by identifying discrepancies.

Table 1

List of clinical data management activities

Data collection
CRF tracking
CRF annotation
Database design
Data entry
Medical coding
Data validation
Discrepancy management
Database lock

Database designing

Databases are the clinical software programmes created to make it easier for CDM to perform several investigations.

[6] In general, these tools are simple to use and built to compliance with regulatory criteria. To ensure data security, "system validation" is carried out, during which system specifications, [7] user requirements, and regulatory compliance are assessed prior to deployment. The database contains information about the study's goals, time intervals, visits, investigators, sites, and patients, and CRF layouts are created for data entry. Before being used for real data gathering, these entry screens are validated using fake data.

Data collection

The CRF, which can be found in a paper or electronic version, is used for data gathering. The typical approach involves using paper CRFs to collect data responses, which are then converted to a database through internal data entry. The investigator completes these paper CRFs in accordance with the completion standards. In the e-CRF-based CDM, the investigator or a designated person will log in to the system and enter data on the spot. Errors are less likely to occur and inconsistencies are resolved more quickly with the e-CRF approach. Many pharmaceutical companies are using e-CRF options as they attempt to shorten the time required for drug development processes by speeding up the processes involved (also called remote data entry).

CRF tracking

The Clinical Research Associate (CRA) will check the CRF entries for accuracy and completeness. The CDM team receives the recovered CRFs. The CDM team will keep track of and record the recovered CRFs. CRFs are manually examined for missing pages and unintelligible data to ensure that the data is not lost. If there is any missing or unclear data, the investigator is contacted for explanation, and the problem is fixed.

Data entry

Data entry is done in accordance with the rules created for the DMP. This only applies to paper CRF that was retrieved from the sites. Double data entry is typically carried out by two operators entering the data independently. [8] The second pass input, or entry made by a second person, helps in reconciliation and verification by pointing out transcription mistakes and discrepancies brought on by unintelligible data. In addition, compared to a single data input, double data entry helps to produce a cleaner database. Double data entry ensures improved consistency with paper CRF, as seen by a lower mistake rate, according to earlier studies. [9]

Data validation

Data validation is the process of ensuring that data is accurate and compliant with protocol requirements. In order to assure the validity of the data entered, edit check programmes are created to detect discrepancies in the data that are contained in the database. The logic conditions listed in the DVP are followed when writing these programmes. These edit check systems are initially tested using fictitious data that contains errors. A data point that fails to pass a validation check is referred to as a discrepancy. Discrepancy may result from incomplete or inconsistent data, range checks, protocol violations, or missing data. Data validation processes are widely used in e-CRF based investigations to find discrepancies. After logging into the system, the investigators will fix these errors. At frequent periods throughout CDM, data processing quality is continuously monitored. For example, if the inclusion criteria state that the patient's age must fall within the range of 18 and 65 (inclusive), an edit programme will be created for the two conditions of age 18 and age >65. A discrepancy will occur if the condition turns out to be TRUE for any patient. The system will signal these inconsistencies, and Data Clarification Forms (DCF) can be produced. Documents called DCFs contain questions about the discovered disparities.

Medical Coding

Medical coding helps identify and correctly classify medical terms associated with clinical trials. An online medical dictionary is used to categorise events. Technically, this activity requires knowledge of medical terminology, understanding of disease entities, drugs used, and basic knowledge of the pathological processes involved. Functionally, knowledge of the structure of electronic medical dictionaries and the hierarchy of classifications available therein is also required. During the study, pre- and on-medication adverse events, and pre-existing or concomitant illnesses were coded using available medical dictionaries. In general, the MedDRA for Regulatory Activities (MedDRA) is used for coding adverse events and other diseases, and the World Health Organization-Enhanced Drug Dictionary (WHO-DDE) is used for coding drugs. These dictionaries contain the correct classes for each classification of adverse events and drugs. Other dictionaries are also available for data management (for example, WHO-ART is a dictionary for side effect terms). Some pharmaceutical companies use custom dictionaries that fit their needs and meet standard operating procedures.

Medical coding helps classify CRF-reported medical terms into a standard dictionary to ensure data

consistency and avoid unnecessary duplication. For example, researchers may use different terminology for the same adverse event, but it is important to code them all in a single standard code while maintaining consistency. Proper coding and classification of adverse events and drugs is important, as incorrect coding can obscure safety issues or highlight false safety concerns associated with the drug.

Database locking

After proper quality checks and assurance, final data validation is performed. If there are no deviations, the SAS data set is finalised in consultation with a statistician. All data management activities must be completed before the database is locked. To ensure this, a pre-lock checklist is used to ensure the completion of all activities. This happens because once the database is locked it cannot be changed in any way. The database is locked and clean data is extracted for statistical analysis when lock approval from all parties has been received. The database cannot normally be modified. However, in the event of a critical problem or other critical operational reason, privileged users can modify data even after the database is locked. This calls for enough rationale for modifying a closed database, as well as good documentation and the maintenance of an audit trail. Data extraction is done after locking from the final database. Here's that archive.

Roles and Responsibilities in CDM

In a CDM team, team members are assigned various roles and responsibilities. Members of the CDM team must have a degree in the life sciences and be proficient in computer programmes. Ideally, medical coders should be medical graduates. However, EMT graduates are also employed in industry as medical coders. There are several important roles for all her CDM teams. The list of roles below can be considered the minimum requirements for a CDM team.

- Data controller
- Database Programmer/Designer
- Medical coders
- Clinical Data Coordinator
- Quality control staff
- Data entry associate

Data controllers are responsible for overseeing the entire CDM process. The data controller prepares the DMP and approves CDM procedures and all internal documents related to her CDM activities. Controlling and assigning database access to her members of the team is also the responsibility of the data manager. A database programmer or designer constructs study databases, conducts CRF annotations, and implements edit checks for data validation. Also, is in

charge of creating the database data input screens and testing adjustments using dummy data. A medical coder will code adverse events, medical history, comorbidities, and concomitant medications administered during the study. The Clinical Data Coordinator drafts the CRF, prepares instructions for completing the CRF, and is responsible for DVP development and discrepancy management. All other CDM-related documents, checklists and guidance documents are produced by the Clinical Data Coordinator. The Quality Officer verifies the correctness of data entry and performs data audits [10]. There may be another Quality Assurance person who validates the data entered. In addition, the quality control officer reviews the documentation of the procedures followed. A data entry clerk tracks receipt of CRF pages and performs data entry into the database.

Conclusion

The need for medication development to be accelerated by pharmaceutical firms and for regulatory agencies to establish quality systems to guarantee the creation of high-quality data for accurate drug evaluation has resulted in the evolution of CDM. There is a steady transition from paper-based to electronic data management systems in order to match expectations. The CDM method and systems have benefited from technology advancements, which has produced encouraging outcomes in terms of data generation speed and quality. Professionals in CDM should simultaneously guarantee that the requirements for enhancing data quality are met. [11] Since that CDM is a specialism unto itself, the systems, procedures, and standards that are being used should be assessed. The establishment of legislation to specify the processes to be followed and the data standards, as well as the standardisation of the data management process across businesses, would be the largest regulatory difficulty. The planning and execution of data management systems in a dynamic operating environment where the quick speed of technological advancement outpaces the current infrastructure would be the largest challenge from the industry's standpoint. Notwithstanding this, CDM is developing into a standard-based clinical research entity by balancing the demands placed on existing systems and their limitations with the demands of commercial and technology advancements.

References

- [1] Gerritsen MG, Sartorius OE, vd Veen FM, Meester GT. Data management in multi-center clinical trials and the role of a nation-wide computer network. A 5 year evaluation. *Proc*

- Annu Symp Comput Appl Med Care*. 1993;659–62. [PMC free article] [PubMed] [Google Scholar]
- [2] Lu Z, Su J. Clinical data management: Current status, challenges, and future directions from industry perspectives. *Open Access J Clin Trials*. 2010; 2:93–105. [Google Scholar]
- [3] CFR - Code of Federal Regulations Title 21 [Internet] *Maryland: Food and Drug Administration*. [Updated 2010 Apr 4; Cited 2011 Mar 1]. Available from: <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=11.10>.
- [4] Study Data Tabulation Model [Internet] *Texas: Clinical Data Interchange Standards Consortium*. c2011. [Updated 2007 Jul; Cited 2011 Mar 1]. Available from: <http://www.cdisc.org/sdtm>.
- [5] CDASH [Internet] *Texas: Clinical Data Interchange Standards Consortium*. c2011. [Updated 2011 Jan; Cited 2011 Mar 1]. Available from: <http://www.cdisc.org/cdash>.
- [6] Fegan GW, Lang TA. Could an open-source clinical trial data-management system be what we have all been looking for? *PLoS Med*. 2008; 5:e6. [PMC free article] [PubMed] [Google Scholar]
- [7] Kuchinke W, Ohmann C, Yang Q, Salas N, Lauritsen J, Gueyffier F, et al. Heterogeneity prevails: The state of clinical trial data management in Europe - results of a survey of ECRIN centres. *Trials*. 2010; 11:79. [PMC free article] [PubMed] [Google Scholar]
- [8] Cummings J, Masten J. Customized dual data entry for computerized data analysis. *Qual Assur*. 1994; 3:300–3. [PubMed] [Google Scholar]
- [9] Reynolds-Haertle RA, McBride R. Single vs. double data entry in CAST. *Control Clin Trials*. 1992; 13:487–94. [PubMed] [Google Scholar]
- [10] Ottevanger PB, Therasse P, van de Velde C, Bernier J, van Krieken H, Grol R, et al. Quality assurance in clinical trials. *Crit Rev Oncol Hematol*. 2003; 47:213–35. [PubMed] [Google Scholar]
- [11] Haux R, Knaup P, Leiner F. On educating about medical data management - the other side of the electronic health record. *Methods Inf Med*. 2007; 46:74–9. [PubMed] [Google Scholar]
- [12] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3326906/>

